

“If You’re Not Keeping Score, You’re Just Practicing”: A Lean Health Care Program Evaluation Framework

February 6, 2012

Martin L. Puterman¹, Yue Zhang², Sema K. Aydede³, Becky Palmer³, Stuart MacLeod³, Hessam
Bavafa⁴ and Jennifer MacKenzie³

¹Sauder School of Business, University of British Columbia, British Columbia, Canada

²College of Business and Innovation, The University of Toledo, Ohio, U.S.A.

³Provincial Health Services Authority, British Columbia, Canada

⁴The Wharton School, University of Pennsylvania, Pennsylvania, U.S.A.

Corresponding author: Martin L. Puterman, Sauder School of Business, University of British
Columbia, 2053 Main Mall, Vancouver, British Columbia, Canada, V6T 1Z2. Tel: 1-604-822-8388;
Fax: 604-822-9574; E-mail: martin.puterman@sauder.ubc.ca

Abstract

Objective. To provide a rigorous framework for evaluating a Lean health care program based on an approach developed to evaluate the Provincial Health Services Authority (PHSA) Lean initiative.

Design. Data for analysis includes performance metrics for Lean activities, system level performance measures, staff interviews and longitudinal surveys and inputs.

Result. This paper describes the essential components of a Lean evaluation including short and long term efficiency gains, quality and safety improvements, staff engagement, resources and financial inputs.

Conclusions. The proposed evaluation framework for Lean interventions, based on a rigorous statistical foundation, considers key performance measurement dimensions and can be widely applied. It addresses a major failure of Lean programs; the lack of formal evaluation.

Key Words. Lean, evaluation, health care, performance metrics, data.

Introduction

Lean methods date back to the Toyota Production System which was developed shortly after World War II [1]. Recently they are finding wide spread usage in health care. In industry, the primary objective of Lean is to enhance corporate earnings, but in health care, lean seeks to improve patient-centered care and system efficiency [2]. Lean evaluations have tended to be ad hoc, usually focusing on before and after pictures, graphical displays or vignettes. With Lean's widespread adoption in health care, a framework for its evaluation is sorely needed.

This paper provides guidelines for evaluating a comprehensive Lean program based on insights we gained developing a framework to evaluate "imPROVE", a significant Lean initiative undertaken by the Provincial Health Services Authority (PHSA) in British Columbia (BC), Canada. This paper will not present results of the imPROVE evaluation but will focus on concepts, methods and observations.

PHSA operates 10 agencies including BC Children's Hospital, BC Women's Hospital and the BC Cancer Agency, employs 8500 people and projects budget expenditures of \$1.76 billion in 2010. PHSA launched imPROVE in 2007 to enhance quality and safety, staff engagement, system sustainability and efficiency. imPROVE now conducts approximately 75 "kaizen" events or "Rapid Process Improvement Workshops" (RPIWs) each year focussing on issues identified through "value stream mapping" at each of its ten agencies. The events use Lean tools to remove inefficiencies and non-value adding activities, or reorganize work space. In collaboration with outside consultants, imPROVE staff lead and facilitate the events with participation from agency staff, management and the executive.

Lean carried out in this way presents challenges for formal evaluation. Rigor suggests controlled studies in which serial changes in measures are compared between randomly selected “treated” institutions and comparable “control” institutions. Such an approach would be suited for a health system or authority where the primary units are hospitals, but PHSA’s Lean implementation was carried out at disparate agencies so that natural comparators were not available. Consequently impacts were analyzed on a serial within-agency basis. Unfortunately this may confound results with effects of other system change initiatives and environmental factors.

Lean and Its Evaluation

Lean has been successfully employed in a variety of manufacturing industries [3, 4, 5]. Because of these successes, there has been an increased interest in implementing Lean in the health care sector in laboratories [6, 7, 8, 9], intensive care units [10], emergency departments [11, 12, 13], operating rooms [15,16], pharmacies [17,18], and hospital-wide [19-22].

Although Lean has been applied widely in the health care sector, formal evaluations have been lacking. Both data collection methods and statistical analyses lack rigor. Further, most studies focus primarily on operational improvements, such as wait time reduction or increase in throughput in a single process and rarely consider system level impact. A few studies report costs and savings. Even fewer discuss changes in organizational culture and engagement. None provide a comprehensive evaluation taking into account all aspects of a Lean initiative.

Therefore, there is an urgent need for a comprehensive evaluation framework that is based on rigorous statistical foundation, considers multiple dimensions and is widely applicable. To the

best of our knowledge, there is no such evaluation framework for Lean interventions in the health care literature. Nonetheless, evaluation frameworks have been developed for other improvement studies. [23] describes a comprehensive evaluation of family nurse practitioner programs that takes into account operational performance, cost, satisfaction and provider attitudes. [24] provides a comprehensive guide to evaluation in public health interventions; it identifies evaluation components and gives guidance for designing an effective process evaluation plan. [25] provides an overview of the use of the Lean management concepts in the Scottish public sector based on literature reviews, cross-case study analysis, surveys and pilot studies. Recently, [26] describes an evaluation framework and approach for a patient safety initiatives by the U.S. based Agency for Health Care Research and Quality.

The Proposed Evaluation Framework

As an initial step in the evaluation process, agreement on terminology and definitions is crucial. In our evaluation of imPROVE, we found that a glossary of common terms focused discussions. A case study that illustrates the concepts below appears in Appendix 2.

Table 1 provides an overview of the evaluation components including short and long term:

- efficiency gains,
- quality and safety improvements,
- staff engagement enhancements and
- financial and resource inputs.

Evaluating Individual RPIW Outcomes

RPIWs, which are intensive sub-process interventions, are the basic building blocks of any Lean program. They seek to:

- Remove waste;
- Achieve value stream goals;
- Train staff in Lean methodology;
- Provide staff with hands on experience using Lean methods.

We identified the following three broad categories of RPIWs, each requiring different performance metrics:

1. flow improvement,
2. defect reduction and
3. workplace reorganization (“Sort, Straighten, Sweep, Schedule and Sustain”, or 5S).

Examples of these RPIW types include reducing operating room turn-over time, improving documentation and centralizing medication information to reduce medication reconciliation errors during patient admission and transfer, and reorganization of bedside carts.

Prior to beginning an RPIW, Lean leaders and area management must clearly define outcome metrics and develop data collection procedures and forms. The RPIW and its performance metrics must be aligned with overall value stream goals. Reliable baseline data or “current state” information must be collected and sample sizes and sampling methods specified to ensure validity. Further, a *primary event metric* must be identified. It must be quantifiable and aligned with value stream goals. The next step is to express RPIW goals in terms of a *target* or “future state” for this primary event metric. The target must be sufficiently “audacious” so that

it cannot be achieved without innovation and the change that it corresponds to be important within the context of the problem under study.

In choosing sample sizes for data collection in RPIWs the following considerations should be noted.

- When data is available from administrative data systems, no sampling is needed.
- When the metric is a rate, such as adverse events per day, complete data should be retained.
- When there are day of the week or week in the month effects, data collection and sampling methods must account for them.
- When the data to compute the metric needs to be sampled and there is sufficient data available, the sample size should be such that the performance metric mean can be estimated accurately. Standard confidence interval formulas provide guidance. When the standard deviation of the measurement is not known, it should be approximated by a simple rule of thumb such as asking the team for a plausible range for the measurements and choosing the standard deviation to be the width of the interval divided by four. (This assumes a 95% prediction interval for the values which is approximately equal to ± 2 standard deviations).
- When the potential sample size is small, collect as many measurements as possible.

For 5S events, a standard scoring system has been well established although events of this type could also impact operational measures that these metrics don't account for.

General guidelines are to retain *all* data, record the date on which it was collected, summarize it in terms of means and standard deviations or medians, quartiles and extremes and represent it graphically in box plots (Figure 1) so that comparisons between metrics at different times can be easily undertaken. Often, in Lean implementations, data reliability and collection rigor are sacrificed for expediency.

The evaluation must distinguish short, intermediate and long term effects of an event. This requires both reliable baseline measurements and adequate follow up data. The primary metric should be evaluated at several points of time prior to the intervention to avoid the impact of unusual events, day of week effects or other biases. Subsequently, the metric should be computed immediately following the intervention for several weeks thereafter. These measurements are primarily to ensure that the intervention impact has occurred and has stabilized. Finally the metric should be monitored and collected , for example, at one month, three months, six months and twelve months post-intervention. This data will indicate whether change has been sustained.

It may be challenging to motivate staff to collect this data, especially after completion of the intervention, but it is extremely important for determining whether changes had lasting impact and whether targets had been attained and maintained. The importance of measurement must be engrained during the RPIW training period; adding data analysts to lean teams can provide considerable value.

Combining RPIW Measures

To measure the overall effect of the Lean program, individual RPIW metrics must be combined. This presents challenges because metrics may be on different scales, the measurements underlying the metrics may vary by different amounts and events may occur at different points of time. We recommend using composite metrics based on:

- number or percentage of events that have achieved their targets,
- percentage of events that have shown no improvement, and
- percentage improvement relative to the difference between baseline and target.

These should be presented graphically; see Figure 2 for an example. These measures can be further subdivided by agency, hospital, type of unit and type of intervention and compared formally using analysis of variance. Note that this might be statistically problematic because of the longitudinal nature of the data. Also results will be sensitive to how the targets are set.

For managerial reporting, event status should be summarized using a “traffic light system”. This requires precise definitions for coding an indicator as red, yellow or green. We recommend that this coding be based on change relative to baseline and the target. A plausible scale is green denotes that the target is attained, yellow that there was an improvement of greater than 50% of the difference between the target and baseline and red denotes any other outcome. This can be further refined to account for variability. For example, green might be refined to indicate that the target is attained and variability has not increased by more than 10%. Again the lights should be presented longitudinally to show short and long term event status.

Evaluating System Level Impacts

Operationally, the primary goal of Lean is to improve system efficiency measured in terms of throughput, lead times and lengths of stay. It is well known in operations management [26] that speeding up *non-bottleneck* resources or sub-processes need not improve overall process flow. Since RPIWs do not always focus on bottlenecks, system level measures are necessary to determine if improvements at the sub-process level impact the system as a whole. Figure 3 describes the relationship between Lean activities (RPIW or value stream) and their metrics.

System level data should be available from administrative data bases and if not, should be added and collected routinely. Even if such data is routinely recorded, obtaining it and identifying it with a particular value stream may present challenges. Often, time stamps may not be aligned with those used in the Lean study. Also, since some RPIWs might only target specific patient types, patient categories must be linked to the data so that those impacted by the Lean activities can be singled out and provide the basis of analysis.

For Lean initiatives focusing on quality of care and patient safety, appropriate measures include degree of compliance with patient care standards, incidence of adverse events, near misses and safety hazards averted. The availability of patient safety data will be dependent on institutional efforts to implement an effective safety reporting system [28,29].

Evaluation of system level impacts must be based on one or two metrics that align with the value stream objectives. Time series regression models are well suited for this analysis but model formulation depends on whether or not there is a control group. In the presence of a control group, one should compare the difference in post-intervention effects between the “treated” and “control” groups; in the absence of a control, pre-and post-intervention changes

become the basis for analysis. This analysis requires between one and two years of pre-intervention data and as much post-intervention data as possible. Since imPROVE usually had more than one intervention in a particular value stream, some different modelling options for the intervention are possible. The time series regression model below is most suitable and flexible in the presence of a control:

$$\text{Metric}_t = \beta_0 + \beta_1 \text{Intervention}_t + \beta_2 \text{Treatment}_t + \beta_3 \text{Intervention}_t * \text{Treatment}_t + \beta_4 \text{Covariate}_t + \text{Error}_t \quad (1)$$

where the subscript t denotes the time period in which the measurements occurred, the Intervention_t variable is coded as 0 in all periods t prior to the completion of the intervention and 1 in all periods after the final intervention is completed, the Treatment_t variable is coded 1 for a site or unit in which an intervention occurred and 0 in a control site and the term Covariate_t is used to adjust for one or more background effects such as varying patient volumes or acuity levels. Because the data is longitudinal, modelling of the error term requires careful analysis. For imPROVE data a time series regression model with AR(1) errors [30] was suitable although more complex error models may be necessary. When there is more than one intervention, several intervention variables can be used but we prefer the use of a single intervention variable since the focus in this phase of the analysis is on how the Lean program impacts the value stream and not on the impacts of the individual interventions.

In the absence of a control, the two terms involving treatment should be omitted from (1) and inference should be based on the significance and magnitude of the coefficient β_1 . When there

is a control, inference should be based on β_3 which measures the difference in the change from pre- to post-intervention between the treatment and control groups.

Patient satisfaction surveys pre- and post-intervention can also provide insights into system level effects. They should use validated instruments to elicit information on patient perceptions of ease and timeliness in accessing care, and its patient-focus and quality. These surveys are usually administered to random samples of patients. Patient survey data may be routinely reported in standard management dashboards, however sub-group analysis might be necessary to align the data with value streams in which there were interventions.

Evaluating Staff Engagement

A high performance health care system requires high levels of staff engagement [31]. Lean strives to achieve this by providing staff with a set of continuous improvement tools. [32] attributes Toyota's Lean success not to its process innovations, but to its ability to create a learning environment where workers were engaged and became agents of continuous improvement. A comprehensive Lean evaluation must explore if these ends are achieved. Interviews and RPIW evaluations can provide some insights on a short term basis, but since cultural change is the over-arching goal, this can best be measured qualitatively through interviews and focus groups and quantitatively through longitudinal surveys and absenteeism, transfers, leave taking and resignations.

Post-RPIW Evaluations and Interviews

At the end of each RPIW, all participants should complete an evaluation to ascertain whether:

- Lean objectives and tools have been effectively communicated,
- RPIW objectives have been attained,
- learning has taken place,
- the experience has been worthwhile, and
- the experience changed the way participants might carry out their future work.

Standardized and validated forms should be used, and a formal scoring system developed which summarizes information on the basis of the dimensions of education, impact and engagement. Information gathered in this way can not only be used to improve the RPIW process but also provide some early feedback on whether an environment has been created that involves staff and motivates them to consider using Lean in their future work. This data is best collected anonymously so that it will be difficult to measure any impact on individual staff. Monitoring scores on a continual basis will enable Lean leadership to continuously improve the quality of the RPIW experience.

Structured and open ended interviews can provide more in depth feedback; Appendix 1 provides our script. Formal analysis using qualitative research methods is possible; however, if the objective is to provide preliminary feedback on the effectiveness of the process, responses can be categorized and summarized with some indication of the frequency of comments. Rigor suggests that it would be best to compare responses of individuals who have been involved in RPIWs to others who have not.

Longitudinal Surveys and Employment Data

Longitudinal staff satisfaction surveys and high level HR data provides more tangible and quantifiable engagement measures. Most organizations periodically administer staff satisfaction surveys. They are usually carried out by opinion research organizations which provide validated and reliable surveys, quantitative scales, rigorous sampling methods, proper statistical analysis and frames of reference for results. While these surveys are not designed for intervention analysis, nonetheless they provide reliable data to investigate the impact of Lean on engagement. While scores on specific questions such as “Last year, I have had opportunities at work to learn and grow” should highly correlate with a successful Lean implementation, other more general questions and summary measures also provide guidance.

HR management reports on absenteeism, sick leaves, leaves of absence and transfers provide another data source. In a large organization, such information should be available as part of dash-boards or score cards. One would hypothesize that if Lean was improving staff engagement, one should see fewer absences, sick leaves, leaves of absence and transfers because of greater levels of job satisfaction.

Of course, even when both sources of engagement data are available, its analysis and attribution to a Lean initiative remains contentious. Ideally one should see improvement in survey results and HR metrics after interventions, but more careful analysis is necessary to associate these with interventions. Fortunately, data of this type is amenable to rigorous statistical analysis using models such as (1) above, which take into account controls, timing and intensity of interventions and background effects. This requires disaggregating data to the unit level and obtaining standard errors.

For any HR metric, comparisons should be between “treated” and “control” units adjusting for the number of interventions and baseline effects. Statistical analyses should account for this effect by either explicitly including the number of interventions or only considering units as “treated” if they received more than a pre-specified number of interventions. Also, in our experience, several of the RPIWs targeted areas with performance challenges so that baseline scores would be low in these units.

Quantifying Inputs and Resources

Numerous inputs are required to implement and sustain a Lean initiative. In a top-down implementation, primary inputs include start up costs, operating costs, and staff and executive time. These must be balanced against the above outcomes to evaluate the overall impact of Lean.

Operating Costs and Salaries

This input component consists of all variable expenses directly related to implementing Lean including consultant expenses, personnel expenses and materials. The other short-term component is staff time for those who take part in Lean activities. This data may be obtained directly through creating a billing code for staff, such as nurses, clerks and technicians who are on the organizational payroll, and from billed costs from professionals who contract for their time. Double counting should be avoided when jobs are backed-filled to replace staff involved in Lean activities. This information can be summarized in terms of staff-hours or actual payments.

Start Up Expenses and Executive Time

Embarking on Lean initiative involves a range of start up costs. These include site visits to other health care organizations, trips to Japan and acquisition and preparation of materials. The amount of executive time is difficult to quantify but a portion of the salary or time spent in formulating and launching Lean should be captured. The use of Lean consultants also represents a major expense.

Further Observations

When evaluating Lean, the target audience will dictate the frequency and detail at which results should be reported. Executive teams might require quarterly reports while boards or public bodies might only require annual reporting. Lean leadership requires continuous monitoring. Lean tools such as “performance walls” might effectively communicate some of this information, especially at the RPIW level, but graphical presentations and tables must be chosen judiciously.

One of the biggest challenges facing Lean evaluators is attaining timely and accurate data at the appropriate level of granularity. Obtaining such data requires strong support from institutional IT leadership. To this end, we recommend including IT personnel and data analysts in the Lean execution team.

While Lean is often an effective vehicle for identifying and improving system performance, several areas such as staff and patient scheduling, and capacity planning are too complex or

inappropriate for Lean solutions. Techniques from operations research including optimization and simulation provide appropriate alternatives.

A case study illustrating these concepts appears in a longer version of this paper that is available from the authors.

Unbiased, objective and preferably at arms length Evaluation is an essential component of any change initiative. It is best done outside of the organization where the Lean initiative is based. University research institutes and health evaluation units often provide such services.

References

1. Ohno T. *Toyota Production System, Beyond Large Scale Production*. Portland: Productivity Press, 1988.
2. Institute of Medicine. *Crossing the Quality Chasm: a new health system for the 21st Century*. Washington, DC: National Academy Press, 2001.
3. Liker, JK, (ed.), *Becoming Lean, Inside Stories of US Manufacturers*, Productivity Press, Portland, Ore., 1998.
4. Melton T. The benefits of lean manufacturing: what lean thinking has to offer the process industries. *Chem Engr Res and Design* 2005; A6: 662–673.
5. Abdulmaleka FA, Rajgopalb J. Analyzing the benefits of lean manufacturing and value stream mapping via simulation: A process sector case study. *Intl J Prod. Eco.* 2007; 1: 223-236.
6. Persoon TJ, Zaleski S, Frerichs J. Improving preanalytic processes using the principles of lean production (Toyota Production System). *Am J Clin Path* 2006; 125: 16-25.
7. Raab SS, Andrew-JaJa C, Condel JL *et al.* Improving Papanicolaou test quality and reducing medical errors by using Toyota production system methods. *Am J Obstet Gynecol* 2006; 194, 57-64.
8. Zarbo, RZ, D'Angelo R. The Henry Ford System: Effective reduction of process and waste in surgical pathology. *Am J Clin Path* 2007; 128: 1015-1022.

9. Hassell, LA, Glass CF, Yip C *et al.* The combined positive impact of Lean methodology and Ventana Symphony autostainer on histology lab workflow. *BMC Clin Path* 2010; 10:2.
10. Shannon, RP, Frndak D, Grunden N *et al.* Using real-time problem solving to eliminate central line infections. *Jt Comm J Qual Patient Saf* 2006; 32: 479-487.
11. King DL, Ben-Tovin DI, Bassham J. Redesigning emergency department patient flows: application of lean thinking to health care. *Emerg Med Australas* 2006; 18: 391–397.
12. Kelly AM, Bryant M, Cox L *et al.* Improving emergency department efficiency by patient streaming to outcomes-based teams. *Aust Health Rev* 2007: 31; 16-21.
13. Decker WW, Stead LG. Application of lean thinking in health care: a role in emergency departments globally. *Int J Emerg Med* 2008; 3, 161-162.
14. Dickson, EW, Anguelov, Z, Vetterick, D, Eller, A., Singh, S. Use of Lean in the Emergency Department: A Case Series of 4 Hospitals. *Ann. Emerg. Med.* 2009: 54, 504-510.
15. Leslie M, Hagood C, Royer A *et al.* Using lean methods to improve OR turnover times. *AORN Journal* 2006; 5: 849-855.
16. Fairbanks CB. Using Six Sigma and Lean methodologies to improve or throughput. *AORN Journal* 2007; 1: 73-82.
17. Hintzen BL, Knoer SJ, Van Dyke CJ *et al.* Effect of lean process improvement techniques on a university hospital inpatient pharmacy. *Am J Health-Syst Ph* 2009; 22: 2042-2047.

18. Hummer J, Daccarett C. Improvement in prescription renewal handling by application of the Lean process. *Nurs Econ* 2009; 27: 197-201.
19. Kim C, Spahlinger D, Kin J *et al.* Lean health care: What can hospitals learn from a world-class automaker? *J Hosp Med* 2006; 3: 191-199.
20. de Koning H, Verver JPS, van den Heuvel J *et al.* Lean Six Sigma in healthcare. *J Healthc Qual* 2006; 2, 4-11.
21. Furman C, Caplan R. Applying the Toyota Production System: using a patient safety alert system to reduce error. *Jt Comm J Qual Patient Saf* 2007; 33: 376-386.
22. Nelson-Peterson DL, Leppa CJ. Creating an environment for caring using lean principles of the Virginia Mason Production System. *J Nurs Adm* 2007; 37: 287-294.
23. Thompson RS, Basden P, Howell LJ. Evaluation of Initial Implementation of an Organized Adult Health Program Employing Family Nurse Practitioners. *Med Care* 1982; 11: 1109-1127.
24. Steckler A, Linnan L. *Process Evaluation for Public Health Interventions and Research*. San Francisco, CA: Jossey Bass, 2002.
25. Radnor Z, Walley P, Stephens A *et al.* *Evaluation of the Lean Approach to Business Management and its Use in the Public Sector*. Scottish Executive Social Research, 2006.
26. Farley DO, Battles JB. Evaluation of the AHRQ patient safety initiative: framework and approach. *Health Serv Res* 2009; 44: 628-645.

27. Goldratt EM, Cox J. *The Goal: A Process of Ongoing Improvement*. New York: North River Press, 1992.
28. Leape LL., Reporting of adverse events. *New Eng J Med* 2002; 20: 1633-1638.
29. Cochrane D, Taylor A, Miller G *et al*. Establishing a provincial patient safety and learning system: pilot project results and lessons learned. *Healthc Q* 2009; 12: 147-153.
30. Brockwell PJ, Davis RA. *Time Series: Theory and Methods*. 2nd ed. New York: Springer, 2009.
31. http://secure.cihi.ca/cihiweb/products/HCPProviders_07_EN_final.pdf (accessed March 17, 2011).
32. Spear S, Bowen HK. Decoding the DNA of the Toyota production system. *Harvard Bus Rev* 1999; 77: 97-106
33. Odegaard, F, Chen, L, Quee, R, Puterman, ML. Improving the Efficiency of Hospital Porter Services, Part 1: Study Objectives and Results. *Journal for Healthcare Quality* 2007; 29, 4-11, 2007.
34. Odegaard, F, Chen, L, Quee, R., Puterman, ML. Improving the Efficiency of Hospital Porter Services, Part 2: Schedule Optimization and Simulation Model Study Objectives and Results. *Journal for Healthcare Quality*: 29, 2007; 12-18.

Appendix 1: Engagement Interview Questions

1. What do you know about imPROVE or RPIW's or Lean Methodology?
2. Have you been a participant in an imPROVE event? What was this experience like for you? Please describe....
 - If you were asked to participate in an imPROVE event again, would you do so and for what reasons? Yes/No
3. Have you been a lead in an imPROVE event? What was this experience like for you? Please describe....
 - If you were asked to lead or participate in an imPROVE event again, would you do so and for what reasons? Yes/No
4. Would you want to be a participant in an imPROVE event? Yes/No
Please share your reasons why/why not?
5. On a scale of 1-5, how important is it to you to be working in an environment that focuses on continuous improvement?

1 = Not important at all	
2 = Somewhat important	
3 = Neutral	
4 = Important	
5 = Highly important	

6. Do you feel imPROVE has contributed to a culture of continuous improvement? Yes/No
Please describe....

7. Would working in this kind of environment impact your decision to continue your employment within BCCH or BCWH? Yes/No
8. Do you feel more satisfied or engaged with your job and workplace as a result of imPROVE? Why or why not?
9. What impact has imPROVE had on your unit or program? (morale, engagement, satisfaction, efficiency, ...)
10. Is there anything else you would like to share with me regarding imPROVE or your engagement and satisfaction with your work environment?

Appendix 2: Lean Evaluation Case Study

As an example of how to apply the evaluation strategy described in the paper, we describe a study (which pre-dated imPROVE) in which the objective was to reduce out-patient wait times for CT scans at a Vancouver General Hospital in Vancouver, BC. Although lean was not explicitly used, portions of this study would have been well suited to lean methods. After developing the value stream map or process flow diagram for the CT scheduling and delivery process, one source of waste or inefficiency identified was starved or blocked CT machines. The consequence of this was low machine utilization and queues for out-patient CT scans. The cause for this was not having a patient available for scanning and a patient being finished with a scan and still remaining in the CT room. One root cause for these delays was a shortage of porters. To this end the study focus shifted to improving performance of the porter service at the hospital (33,34). For the remainder of the paper we will use concepts from the porter system improvement study to illustrate our recommendations.

Developing Performance Metrics: If we regard the porter system as the value stream under consideration, then system level performance measures should be related to the times from job request to job completion. One potential system level metric is the average time for completion of a task. Overall averages do not tell the whole story because of considerable between task variability and different task urgency levels (STAT, ASAP, Routine) with different target completion times. We chose the system level metric to be the service level which we defined as the percentage of jobs of each category that were completed in the target time.

One potential intervention, which could have been the subject of an RPIW if the lean methodology was being used, was to improve the communication between the porter and the requesting unit so that unit was aware of the forecasted time at which the porter would arrive. This forecasted time could be determined with some degree of accuracy since once an available porter was assigned to a job, the only source of variability was travel time. This would allow unit personnel ample time to get the patient ready for the move, thus avoiding delays in preparing the patient, obtaining documentation and having all necessary equipment for the move, such as wheelchairs or poles, ready. Another potential metric could be based on the time from porter arrival to the unit to pick up a patient to the time the porter leaves the unit with the patient and all equipment and documentation needed for the move.

The above metric, time from arrival to the unit to time up departure from the unit needs to be made precise and a process to collect this data is required. Since such time stamps are rarely available from administrative systems, either the porter or an analyst needs to record this information. At baseline, the porter was “shadowed” by a study analyst (in this case a student) and precise time epochs for recording were determined. These times were when the porter informed the unit clerk that he or she had arrived at the unit and when the patient and all equipment were obtained and the porter departed from the ward. If forms or an automated data collection system was available, all data should be collected; otherwise sampling should be used.

To determine the impact of all porter system improvement interventions, all data on time of request, time of job completion with indicators of urgency, the origin of the move, the

destination of the move and equipment required should be recorded. Hopefully adequate baseline (pre-intervention data) is available. The system level performance metric was chosen to be the percentage of jobs of each time that met the target each month. Or the average time per request type could instead be analyzed. In this study no obvious control was available but if the porter system had undergone lean analysis at one hospital and not at a comparable hospital, the second hospital could serve as a control. In absence of a control the statistical model (1) without the treatment terms should be considered. A potential covariate is monthly call volume.

Table 1: An overview of the Lean evaluation framework illustrating the evaluation dimensions, and the specific items used to measure short- and long-term outcomes. They are discussed in detail in the text.

Time Frame \ Dimension	Efficiency and Sustainability	Quality and Safety	Staff Engagement	Resources and Inputs
Short Term	RPIW Outcomes	RPIW Outcomes	Interviews and post-RPIW evaluations	Materials, Program Support, Staff Salaries and Staff Time
Long Term	System Level Performance Measures and Key Performance Indicators (KPIs)	System Level Performance Measures / KPIs	Longitudinal surveys Absenteeism and leave taking	Start-up Costs and Executive Support

Figure 1: A sample graphical representation showing the evolution of an RPIW primary event metric over time. The horizontal line at 40 represents the target and the circles represent outliers. Note that sample sizes are also reported.

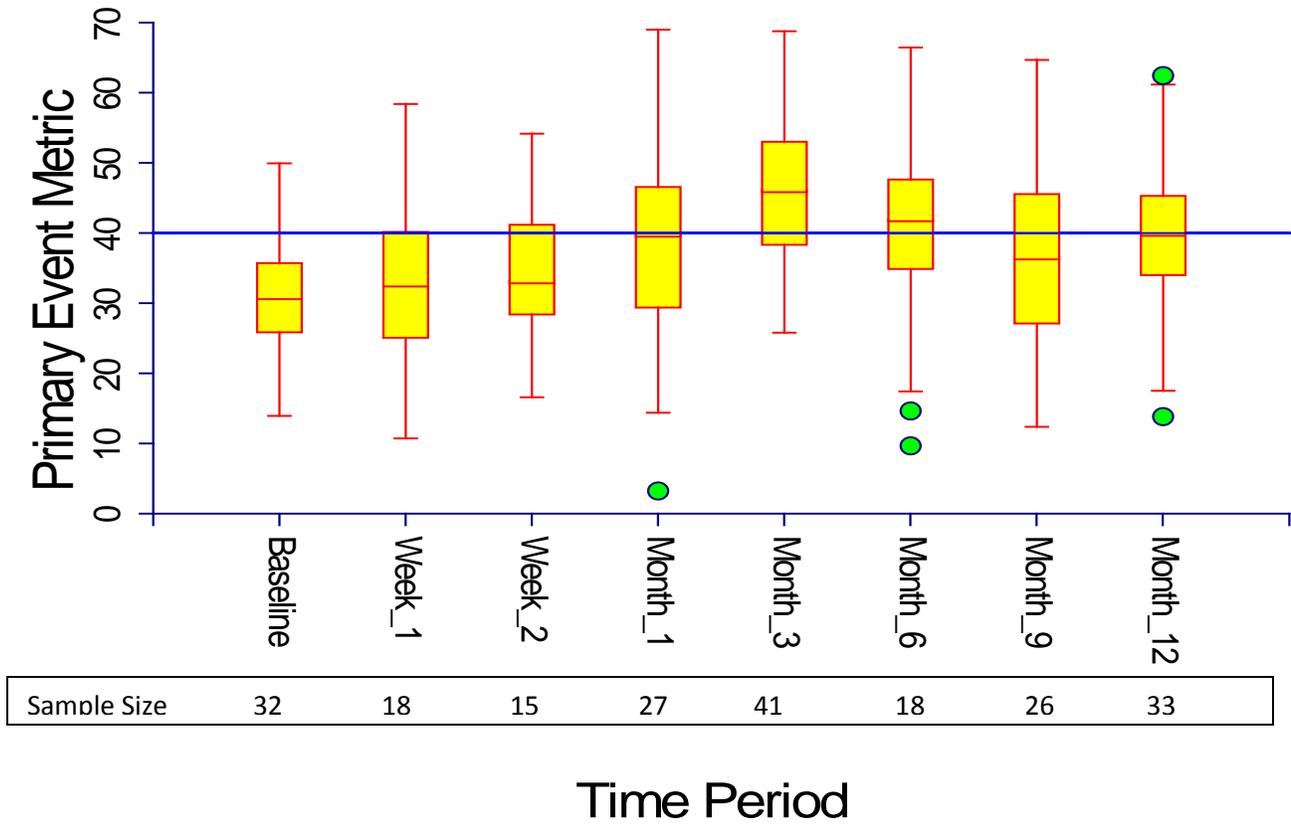


Figure 2: An example graph showing the percentage of improvement relative to the difference between the baseline and target at several time points aggregated across several RPIWs.

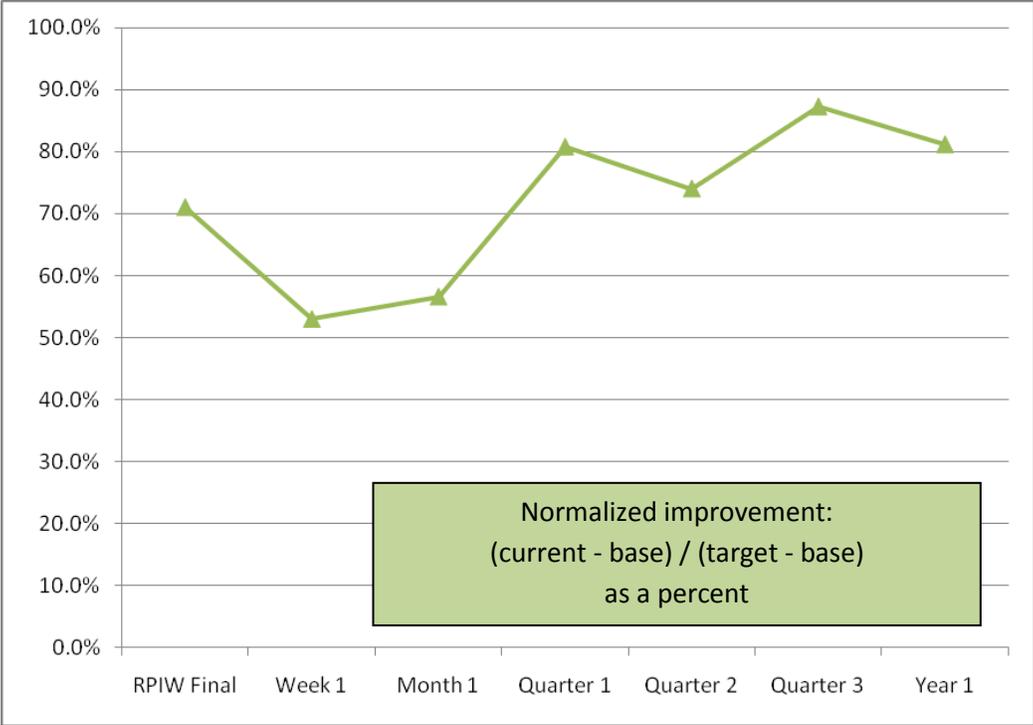


Figure 3; Relationship between lean activities (left) and metric type (right). Dashed arrows suggest potential influences while solid arrows indicate data drivers.

